Framework para Auxílio a Deficientes Visuais utilizando Redes Neurais Convolucionais como Extratoras de Características

Jefferson Antonio Ribeiro Passerini,
Inst.Geociências e Ciências Exatas-IGCE
Universidade Estadual Paulista - UNESP
Rio Claro - SP
Jefferson.passerini@unesp.br

Fabricio Aparecido Breve,
Inst.Geociências e Ciências Exatas-IGCE
Universidade Estadual Paulista - UNESP
Rio Claro - SP
fabricio@rc.unesp.br

***Resumo*—Segundo a ONU existem 2,2 bilhões de pessoas com alguma deficiência visual. Os recursos de visão computacional evoluíram na última década a partir da utilização de CNNs. Este trabalho tem por objetivo estabelecer um framework para detecção de obstáculos como complemento aos recursos existentes para o auxílio aos deficientes visuais. Realizou-se a extração de características através de CNNs com *transfer learning* e aplicou-se diversos classificadores durante os testes, comparando-se ao desempenho de uma CNN treinada. Obteve-se uma mediana de acurácia de 0,9559 no modelo proposto e verificou-se que não existiu diferenças estatísticas relevantes entre as abordagens estudadas para a abordagem de referência.**

Área: Inteligência Computacional.

I. INTRODUÇÃO

Pessoas com deficiências visuais ou totalmente cegas são um aspecto importante de nossa sociedade, o aumento de doenças relacionadas aos olhos e a redução da visão estão representando cada vez mais um desafio a instituições e governos.

Segundo a ONU [1] no mundo existem pelo menos 2,2 bilhões de pessoas com algum tipo de deficiência visual. No Brasil, como levantado durante o censo demográfico em 2010 pelo IBGE [2], existiam 35 milhões de pessoas com algum acometimento visual, onde 506.377 destas pessoas eram totalmente cegas e mais de 6 milhões possuíam grandes dificuldades para enxergar.

As tecnologias assistivas provem a segurança, mobilidade e acesso as pessoas com deficiências [3], assim as pesquisas na área de reabilitação estão avançando e os dispositivos móveis e os aplicativos representam um dos vetores de melhoria da qualidade de vida dessas pessoas [4].

A área de visão computacional com foco no reconhecimento de objetos e detecção de obstáculos obtiveram grandes avanços na última década [5][6], muitos desses sistemas são caros, pesados ou ainda não estão amplamente disponíveis à população [7]. Existem ainda sistemas de acessibilidade que necessitam da utilização de redes de dados para acesso a servidores remotos [8], deste modo o dispositivo mais utilizando atualmente na detecção de obstáculos ainda é a bengala.

Esta pesquisa tem por objetivo desenvolver um framework para a detecção de obstáculos utilizando o *smartphone* como complemento às ferramentas utilizadas pelo deficiente visual em sua locomoção.

II. CONCEITOS E TÉCNICAS

O advento das redes neurais convolucionais (CNNs) determinou um grande avanço na detecção de objetos e classificação de imagens [9][10]. Como qualquer modelo de aprendizado supervisionado as CNNs necessitam de um grande volume de amostras rotuladas em sua fase de treinamento e o processo de preparar amostras rotuladas é muito custoso e demorado [11].

O treinamento é a etapa que exige o maior custo computacional na utilização de CNNs, uma vez treinadas é possível a utilização de modelos como VGG19 [12] ou o InceptionV3 [13] para determinar a existência de obstáculos a partir da captura de imagens em um *smartphone*.

A utilização de transferência de aprendizado (*transfer learning*) permite aplicar camadas que foram treinadas em grandes conjuntos de imagens otimizando uma CNN para aplicações onde os conjuntos de treinamento são escassos [11]. Essa técnica foi utilizada com sucesso em diversas situações [12] [13][14].

Devido ao grande volume de dados extraídos pelas CNN serão aplicados métodos de redução de dimensionalidade como PCA [16] e UMAP [17] e/ou métodos de seleção como Relief-F [18]. Para a classificação priorizou-se a aplicação de classificadores supervisionados.

III. METODOLOGIA DE DESENVOLVIMENTO

O modelo proposto terá algumas etapas como a extração de características, a redução de dimensionalidade e/ou seleção de características a partir dos dados extraídos, e a etapa de classificação. Para comparação de resultados foram treinadas CNNs e aplicadas ao mesmo conjunto de imagens.

Foram treinadas as CNNs Xception, VGG16/19, ResNet 50/101/152, ResNet 50v2/101v2/152v2, Inception V3, InceptionResNet V2, MobileNet e MobileNet V2, DenseNet 121/169/201, NasNetMobile e EfficientNet B0/B1/B2/B3 e B4. Para o modelo proposto, além das CNN citadas, foram utilizadas na extração de características as CNNs EfficientNet B5/B6 e B7.

Os dados extraídos foram expostos aos classificadores Decision Tree, SVM(RBF), SVM(linear), MLP, logistic, Random Forest, Adaboost, Gaussian Naïve Bayes. Os dados passaram por redução de dimensionalidade (PCA/UMAP), ou seleção de características (Relief-F), com vetores de saída variando de tamanhos: 2, 10, 20, 30, 40, 50, 75, 100, 150, 200, 250 e 300. Testou-se também a exposição de todos os dados extraídos aos classificadores (*full*) (Fig.1).

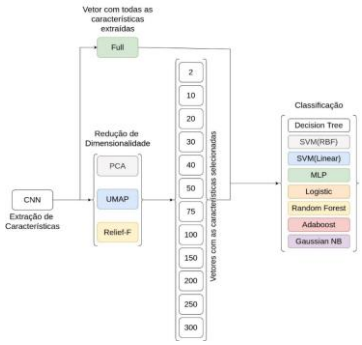

Fig. 1. Metodologia Proposta (Elaborada pelos autores).

A base de imagens para teste foi proposta por eles sendo composta por 342 imagens sendo 175 demonstram caminho livre e 167 determinam obstáculos [19].

IV. Resultados preliminares

Os resultados obtidos nesta fase do trabalho demonstraram que é possível obter resultados semelhantes à classificação utilizando as CNNs treinadas para o objetivo específico do trabalho proposto. A abordagem de referência obteve uma mediana de 0,9412 com um intervalo interquartil de 0,0298 utilizando uma CNN DenseNet201. Comparando a combinações de técnicas selecionadas obteve-se medianas e intervalos interquartis respectivamente 0,9412 e 0,0588 para EfficientNet B0+MobileNet *Full* SVM(linear), 0,9421 e 0,0588 para MobileNet PCA SVM(RBF), 0,9559 e 0,0447 para MobileNet+ResNet50 PCA SVM(linear) e 0,9278 e 0,0663 para MobileNet+ResNet50 PCA SVM(RBF).

Aplicando-se o teste de Kruskal-Wallis obteve-se o resultado [X^2 (4) = 1,609; p=0,807], deste modo verifica-se que não existem diferenças estatísticas relevantes entre os desempenhos dos modelos selecionados.

V. Considerações finais

A partir dos testes aplicados observou-se ser possível obter desempenhos semelhantes a uma CNN treinada. Existe a necessidade do aprofundamento dos estudos devido a incoerências encontradas nos pontos de relevância nas imagens determinados pelas CNNs e que se observou não corresponder aos obstáculos o que pode ser causa de falsa classificação positiva ou negativa.

Referências

[1] ONU. World Report on vision. Disponível em: https://www.who.int/publications/i/item/world-report-on-vision. ONU, Genebra, Suíça, 2019.

[2] IBGE, BRASIL. Censo Demográfico 2010. Instituto Brasileiro de Geografia e Estatística. Brasilia-DF, 2010.

[3] Manduchi, R; Coughlan, J. *(Computer) vision without sight. Comm. ACM*, vol. 55, nº 1, pags 96–104, 2012.

[4] Terven, J.R; Salas, J; Raducanu, B. New Opportunities for Computer Vision-Based Assistive Technology Systems for the Visually Impaired. Computer, vol. 47, nº 4, pp. 52-58, 2014.

[5] Rizzo, J.R.; Pan, Y.; Hudson, T.; Wong, E.K.; Fang, Y. Sensor fusion ecologically valid obstacle identification: Building a comprehensive assistive technology platform for the visually impaired. In 2017 7th ICMSAO. Pag. 1-5. IEEE, 2017.

[6] Lakde, C.K.; Prasad, P.S. Review paper on navigation system for visually impaired people. International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, nº 1, 2015.

[7] Hoang, V.N.; Nguyen, T.H.; Le, T.L.; Tran, T.H.; Vuong, T.P.; Vuillerme, N. Obstacle detection and warning system for visually impaired people on electrode matrix and mobile Kinect. Vietnam Journal of Computer Science, vol. 4, nº 2, pag. 71-83, 2017.

[8] Jiang, B.; Yang, J.; Lv, Z.; Song, H. Wearable vision assistance system based on binocular sensors for visually impaired users. IEEE Internet of Things Journal, vol. 6, nº 2, pag. 1375-1383, 2019.

[9] Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. *Mobilinets: Efficient convolutional neural networks for mobile vision applications*, 2017.

[10] LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature, vol. 521, nº 7553, pag. 436, 2015.

[11] Oquab, M.; Bottou, L.; Laptev, I.; Sivic, J. Learning and transferring mid-level image representative using convolutional neural networks. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June, 2014.

[12] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. Computational and Biological Learning Society, 2015, pag. 1-14.

[13] Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In The IEEE Conference on Computer Vision and Pattern Recognition (CPVR), Jun. 2016, pag.2818-2826, 2016.

[14] Gopalakrishnam, K.; Khaitan, S.K.; Choudhary, A.; Agrawal, A. Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. Construction and Building Materials, vol. 157, pag. 322-330, 2017.

[15] Breve, F.; Zhao, L.; Quiles, M.; Pedrycz, W.; Liu, J. Particle competition and cooperation in networks for semi-supervised learning. IEEE Transactions on Knowledge and Data Engineering, vol. 24, nº 9, pag. 1686-1698, set. 2012.

[16] Jolliffe, I. Springer-Verlag. Principal Components Analysis. Springer Series in Statistics. Springer, 2002.

[17] McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Disponível em: arxiv.org/abs/1802.03426v3, 2020.

[18] Kononenko, I. Estimating attributes: analysis and extensions of RELIEF. **Machine Learning**: ECML-94 Lecture notes in Computer Sciences, v. 784, p. 171-182, Springer, 1 May 1994.

[19] Breve, F.; Fischer, C.N. Visually Impaired Aid using Convolutional Neural Networks, Transfer Learning, and Particle Competition and Cooperation. In International Joint Conference on Neural Networks (IJCNN), mar. 2020.